Published online 9 October 2017 *Nucleic Acids Research, 2017, Vol. 45, No. 22* **12963–12973**
doi: 10.1093/nar/gkx903

Polycytidylation of mitochondrial mRNAs in *Chlamydomonas reinhardtii*

Thalia Salinas-Giegé[1,*], Marina Cavaiuolo[2], Valérie Cognat[1], Elodie Ubrig[1], Claire Remacle[3], Anne-Marie Duchêne[1], Olivier Vallon[2,*] and Laurence Maréchal-Drouard[1,*]

[1]Institut de biologie moléculaire des plantes, CNRS, Université de Strasbourg, 67084 Strasbourg, France, [2]UMR 7141, CNRS, Université Pierre et Marie Curie, Institut de Biologie Physico-Chimique, 75005 Paris, France and [3]Génétique et Physiologie des microalgues, Department of Life Sciences, Institute of Botany, B22, University of Liege, B-4000 Liege, Belgium

Received July 21, 2017; Revised September 18, 2017; Editorial Decision September 25, 2017; Accepted September 25, 2017

ABSTRACT

The unicellular photosynthetic organism, *Chlamydomonas reinhardtii*, represents a powerful model to study mitochondrial gene expression. Here, we show that the 5′- and 3′-extremities of the eight *Chlamydomonas* mitochondrial mRNAs present two unusual characteristics. First, all mRNAs start primarily at the AUG initiation codon of the coding sequence which is often marked by a cluster of small RNAs. Second, unusual tails are added post-transcriptionally at the 3′-extremity of all mRNAs. The nucleotide composition of the tails is distinct from that described in any other systems and can be partitioned between A/U-rich tails, predominantly composed of Adenosine and Uridine, and C-rich tails composed mostly of Cytidine. Based on 3′ RACE experiments, 22% of mRNAs present C-rich tails, some of them composed of up to 20 consecutive Cs. Polycytidylation is specific to mitochondria and occurs primarily on mRNAs. This unprecedented post-transcriptional modification seems to be a specific feature of the *Chlorophyceae* class of green algae and points out the existence of novel strategies in mitochondrial gene expression.

INTRODUCTION

Mitochondria, organelles with a bacterial origin, have preserved a remnant ancestral genome. In all organisms studied so far, mitochondrial (mt) genomes code for only a very small number of proteins. For instance, 32 protein genes are encoded in the mt DNA of the land plant *Arabidopsis thaliana*, 13 in human and only 8 in the yeast *Saccharomyces cerevisiae*. Most of these proteins are subunits of the respiratory chain or the adenosine triphosphate (ATP) synthase and their expression is essential. Despite the common prokaryotic origin of mitochondria, mechanisms allowing mt gene expression have diverged. Studies on mt gene expression in various organisms have highlighted the acquisition of a number of new features and this, in a species-specific manner (1–3). In plant mitochondria, post-transcriptional processes including RNA editing, splicing of introns, maturation of 5′- and 3′-ends of RNA transcripts and RNA degradation play an important role in gene expression (e.g. (4)). Still, many questions remain unresolved (5), for example concerning the identity of promoters, the mechanisms of transcript processing and of translation initiation.

The green alga *Chlamydomonas reinhardtii* is a prime model organism for photosynthesis and flagellar motility (6), but it is also the only photosynthetic organism where mt transformation is possible (7). Mutants impaired in mt respiration are viable in photoautotrophic conditions (8). The *Chlamydomonas* mt genome is linear, very compact, with short intergenic sequences and no introns (9). It codes for only eight proteins, three transfer RNAs (tRNAs) and the small (SSU) and large (LSU) ribosomal RNA (rRNA) subunits fragmented into numerous 'modules' (10), (Figure 1). Transcription starts in the short intergenic region between *nad5* and *cox1* genes and generates two divergent primary co-transcripts which are subsequently processed to generate mature RNAs (11,12). Previous analysis on mt mRNAs showed that their size is close to that of their coding sequence (CDS) indicating the existence of very short 5′- and 3′-UTRs (untranslated regions). This was confirmed for a few transcripts using S1 nuclease protection and primer extension experiments (13,14).

In this work, the *C. reinhardtii* mt transcripts were analyzed using different methods (Supplementary Figure S1A): 5′ and 3′ RACE (Rapid Amplification of

*To whom correspondence should be addressed. Tel: +33 3 67 15 53 98; Fax: + 33 3 67 15 53 00; Email: laurence.drouard@ibmp-cnrs.unistra.fr
Correspondence may also be addressed to Thalia Salinas-Giegé. Tel: +33 3 67 15 53 82; Fax: + 33 3 67 15 53 00; Email: thalia.salinas@ibmp-cnrs.unistra.fr
Correspondence may also be addressed to Olivier Vallon. Tel: +33 1 58 41 50 58; Fax: +33 1 58 41 50 22; Email: ovallon@ibpc.fr

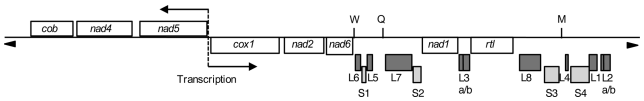

Figure 1. *Chlamydomonas reinhardtii* mitochondrial genome. Schematic map of the *C. reinhardtii* mitochondrial genome. White boxes represent protein-coding genes: (*cob*) apocytochrome *b* of complex III; (*nad1*, *2*, *4*, *5* and *6*) subunits of complex I; (*cox1*) subunit 1 of complex IV; (*rtl*) reverse transcriptase-like protein. The 3 tRNA genes are indicated with the one letter code (W, Q and M). The dark gray and light gray boxes correspond to the LSU (L1–L8) and SSU (S1–S4) RNAs fragments respectively. The bi-directional origin of transcription between *nad5* and *cox1* genes is represented by a dashed vertical line and two horizontal arrows. Telomeric regions are indicated by arrowheads.

cDNA Ends), cRT-PCR (circular Reverse Transcription-Polymerase Chain Reaction) and three protocols based on Illumina single-end sequencing, namely sRNA-Seq (small RNA sequencing, TruSeq), directional WTSS (Whole Transcriptome Shotgun Sequencing, stranded TruSeq) and bi-directional WTSS (non-stranded TruSeq, public data retrieved from SRA). Our results shed light on two singular characteristics of the mRNA transcripts. First, all of them start at the AUG initiation codon of the CDS. Second, they carry post-transcriptionally added C-rich tails at the 3′-end of their short 3′-UTR. This unprecedented observation seems to be a specific feature of the *Chlorophyceae* class of green algae.

MATERIALS AND METHODS

Algae strains and growth conditions

Chlamydomonas reinhardtii strains CC-4351 (*cw15–325 arg7–8* mt+) and CC-5101 (T222 *nit1 nit2* mt+), *Chlorella sorokiniana* 211–32, *Tetradesmus* (*Scenedesmus*) *obliquus* 276–10 and *Polytomella parva* SAG 198.80 were grown on Tris-Acetate Phosphate (TAP) medium (15) supplemented with arginine (100 μg/ml) for CC-4351. *Coccomyxa subellipsoidea C-16* was grown on Modified Bold's Basal Medium (16). Culture conditions were 25°C under white light (10–50 μE m^{-2} s^{-1}). The 'Centre national de ressources biologiques marines' (EMBRC France) provided us with *Chondrus crispus*. The Scandinavian Culture Collection of Algae and Protozoa (SCCAP) provided us with *Cyanophora paradoxa* (K-0262) and *Pedinomonas minor* (K-0264) strains. *Physcomitrella patens* Grandsen strain was provided by R. Resky (University of Freiburg, Germany). *Ostreoccocus tauri* OTTH0595 strain was provided by H. Moreau (Integrative Biology of Marine Organisms CNRS–UMR 7237). For all the organisms, strains and genome accessions are in Supplementary Table S1.

RNA analysis

Crude mt fractions were isolated from cell wall-less *C. reinhardtii* strain CC-4351 by digitonin treatment according to (17). *Chlamydomonas reinhardtii* RNA was prepared from mt fraction or whole cells using TRI Reagent® (Molecular Research Center) according to manufacturer's instructions. For *C. crispus*, *C. paradoxa*, *P. minor* and *P. patens*, RNA was prepared from whole cells using TRI Reagent®. For *C. subellipsoidea*, *C. sorokiniana*, *S. obliquus* and *P. parva*, RNA was prepared from whole cells according to (18). For *C. reinhardtii*, cRT-PCR, 5′ and 3′ RACE experiments were performed with mt RNA, Illumina sequencing with whole-cell RNA. For the phylogenetic study of mt mRNA polycytidylation, 3′ RACE experiments were performed with whole-cell RNA from all species including *C. reinhardtii*.

Oligonucleotides are described in Supplementary Table S2. The cRT-PCR, 5′ and 3′ RACE analyses were performed using Sanger sequencing as in (19). Briefly, 3 μg of mt RNA or 4–6 μg of total RNA were ligated at the 3′-end to adaptor and cDNA was synthesized with SuperScript IV (Invitrogen) primed by the complementary RT primer. PCR was performed with gene-specific primers and RT primers. For some species, nested PCR were necessary to avoid non-specific amplifications. For 5′-RACE, the first-strand cDNA synthesized above was C-tailed with the terminal deoxynucleotidyl-transferase (Invitrogen) in the presence of dCTP. PCR was performed with a gene-specific forward primer and oligo(dG). For cRT-PCR, 4–6 μg of total RNA was circularized by T4 RNA ligase (New England Biolabs) prior to reverse transcription and PCR using gene-specific oligonucleotides. For RT-PCR analysis, 4 μg total RNA treated with DNAse RQ1 (Promega) was reverse transcribed in a 40 μl reaction, using 2,5 μM oligo(dG) primer or a mix of 2.5 μM oligo(dT) and 12 ng/μl random hexamer primer (Promega).

Sequence alignments were generated with MacVector software as described in more details in Supplementary Figure S1B. In order to monitor possible artifacts, 1 ng of synthetic transcript was incubated with 3,5 μg of total RNA and 3′ RACE analysis was performed. Analysis of this synthetic transcript showed that 1–2 additional nucleotides are observed at the 3′ extremity and in very few cases more than 3 nt (Supplementary Figure S1C). For 3′ RACE and cRT-PCR analysis, sequences with more than 3 nt added were therefore considered as *bona fide* 3′-tails.

Illumina sequencing

Single-end 50 nt directional WTSS and sRNA-Seq Illumina datasets (SRA accessions, resp.: SRX2725800 to SRX2725809 and SRX2725861 to SRX2725864) were used, along with ninety bi-directional WTSS datasets (single-end 100 nt) retrieved from SRA as described in (20). WTSS reads were mapped to the mt genome (EU306622) using bwa (mem). sRNA-Seq reads were mapped with either bwa (aln) or Bowtie2 to allow soft-clipping (21,22). Analysis of the mappings used bamtools, bedtools and the IGV browser (23–25).

Phylogenetic and prediction analysis

The maximum likelihood phylogenetic tree of Chlorophyta Class-II NTRs (RNA-nucleotidyl transferases) was generated with PhyML (http://www.atgc-montpellier.fr/phyml), (26) using a clustalW alignment modified to align catalytic residues and trimmed to the conserved regions. Organellar predictions were determined using TargetP (http://www.cbs.dtu.dk/services/TargetP/), (27) and Predalgo (https://giavap-genomes.ibpc.fr/cgi-bin/predalgodb.perl?page=main), (28).

RESULTS

Reannotation of the Chlamydomonas mt genome

Combining various approaches, the mature 5′- and 3′-ends of all mt transcripts of *C. reinhardtii* were precisely determined. Our genome browser (http://chlamy-organelles.ibpc.fr) presents the data as well as a novel annotation that are proposed. Compared to the previous one (EU306622), one mis-sense error in the *cob* CDS is corrected, the start of *nad2* is moved three codons upstream, we modify the ends of almost all rRNAs and reintroduce *rrnL3a* and *rrnL2b* present in the initial annotation (NC_001638). Therefore, the *C. reinhardtii* mt genome encodes 8 mRNAs, 3 tRNAs and 14 rRNA fragments (4 for the SSU, 10 for the LSU), (Figure 1).

Intercistronic maturation generates adjacent 3′- and 5′-termini for almost all RNAs

Concerning transcript processing, the near-absence of WTSS reads connecting successive cistrons indicates that the two primary transcripts that originate in the *cox1_nad5* intergenic region are very efficiently processed into monocistronic RNAs (Figure 2). Gray *et al*. (12) previously showed that *nad5, nad2* and *cox1* mRNAs lack 5′-UTR. Here, 5′ RACE experiments carried out on all mRNAs extend this analysis and indicate that almost all mRNA molecules start at the A of the AUG start codon, meaning that they do not have a 5′-UTR (Figure 2A). In addition, the 5′-end of all mRNAs (except *nad1* and *cob*) is marked by a cluster of sRNAs starting at the AUG initiation codon. A similar result was observed for most rRNAs, (Supplementary Figure S2). It is unknown whether these sRNAs result from protection by RNA-binding proteins, as observed in chloroplasts, or reveal an intrinsic stability of 5′-ends toward exonucleases (20,29). Their length varies greatly (the mode ranges from 16 to 33 nt) between mRNAs (Supplementary Figure S2) and at each locus heterogeneity is high compared to the cosRNAs (clustered organellar sRNAs) described in Arabidopsis organelles (29) and Chlamydomonas chloroplast (20). This suggests that they do not represent footprints of stalled ribosomes or 30S subunits. Whatever the origin of these sRNAs, their presence proved useful to map 5′-ends.

In contrast, all mRNAs had 3′-UTRs (Figure 2B) with sizes corresponding to the downstream intergenic regions (between 14 and 275 nt). The 3′-ends of *nad4*, *nad5*, *cox1*, *nad2*, *nad6*, *nad1* mRNAs were mapped immediately upstream of the next stable transcript, or shortly before. This was true regardless of the nature of downstream genes, coding for a protein, a tRNA or a rRNA. Heterogeneity was slightly higher than at the 5′-end, suggesting that some molecules undergo trimming at the 3′-end after processing at the AUG. The *cob* mRNA, last in the leftward transcription unit and thus not followed by a gene, also harbors a 3′-UTR that ends just before the left telomere. The *rtl* mRNA has two major 3′-ends regions, shortly downstream of the stop codon and just upstream of *rrnL8*. For rRNAs as well, the main rule was that the mature 3′-ends lie at, or just a few nt upstream of the start of the next gene (Supplementary Figures S2 and 3). In addition to the previously described spacer regions separating rRNAs and tRNAs (30,31), a few other short sequences (sometimes as short as 3 nt) were found downstream tRNAs or rRNAs for which no transcripts have been observed (see bidirectional WTSS on genome browser: http://chlamy-organelles.ibpc.fr). In summary, our data show that the 3′-end of a mature RNA most of the time corresponds to the 5′-end of the downstream RNA.

A new type of tails at the 3′-end of mitochondrial transcripts

Sequencing also revealed unusual post-transcriptionally added tails at the 3′ extremities of all mRNAs (Supplementary Figure S3 and Figure 3). These tails are present in high proportions i.e. 35% of total transcripts analyzed using 3′ RACE. Interestingly, the nucleotide composition of these tails is distinct from that described in any other system: while containing certain proportions of Adenosine (20.1%) and Uridine (15.4%), they were strongly enriched in Cytidine (63.7%), a nucleotide hitherto not described in transcript tails. Figure 4 presents randomly selected examples of 3′-tails observed using 3′ RACE for all eight mRNAs. The tails have been partitioned between those predominantly composed of A and U and those composed mostly of C (Figure 3). Both types are equally represented for *nad2* and the A/U-rich tails predominated for *rtl*, but for the six other mRNAs, the C-rich tails are by far the most abundant. Polycytidylation ratios (the fraction of mature mRNAs showing a C-rich tail) based on 3′ RACE clones vary between 3.2–46.8% among genes, with an average of 21.7%. The highest ratios were observed for *nad4* and *cox1*. The length of the tails predominantly fell between 10–20 nt, with an average of 14, a maximum of 29 nt and substantial variations between mRNAs (Figure 5A). In some cases, homopolymeric tails comprising up to 20 consecutive cytidines were observed. Many ‘composite’ tails were also observed, defined here as tails starting with an A/U-rich sequence and ending in a C-rich sequence. For most genes, mRNAs with a complete 3′-UTR, (i.e. ending close to the start of the next transcript) were more likely to show C-rich tails than the truncated ones, where tails were fewer and shorter (Supplementary Figures S3 and 4). The analysis of WTSS data confirmed the presence of C-rich tails on mt mRNAs (Figure 5B–D). C prevailed at almost all positions of the tails (e.g. Figure 5C), even more than with other methods (Supplementary Figure S4), with a smaller proportion of A/U-rich and composite tails. WTSS provides deep coverage, but because it involves priming the reverse transcriptase with degenerate oligonucleotides, it rarely reaches the 3′-end and it is biased toward G/C rich sequences. The ligation-based methods (3′ RACE, cRT-PCR) were therefore favored to describe the composition and length of the tails.

Because 3′-A-rich tails within transcripts are associated with their degradation in bacteria and chloroplasts (32), polycytidylation sites were searched inside transcripts. For this, a *cox1* specific primer positioned far-upstream *cox1* 3′ end was used in 3′ RACE experiments. This analysis identified 61 transcripts truncated within the CDS, but none carried a C-rich tail, (Supplementary Figure S5A). WTSS revealed the presence of 3′-tails within the CDS of *cox1* and *cob*, but the density was 33 and 805 times lower than over

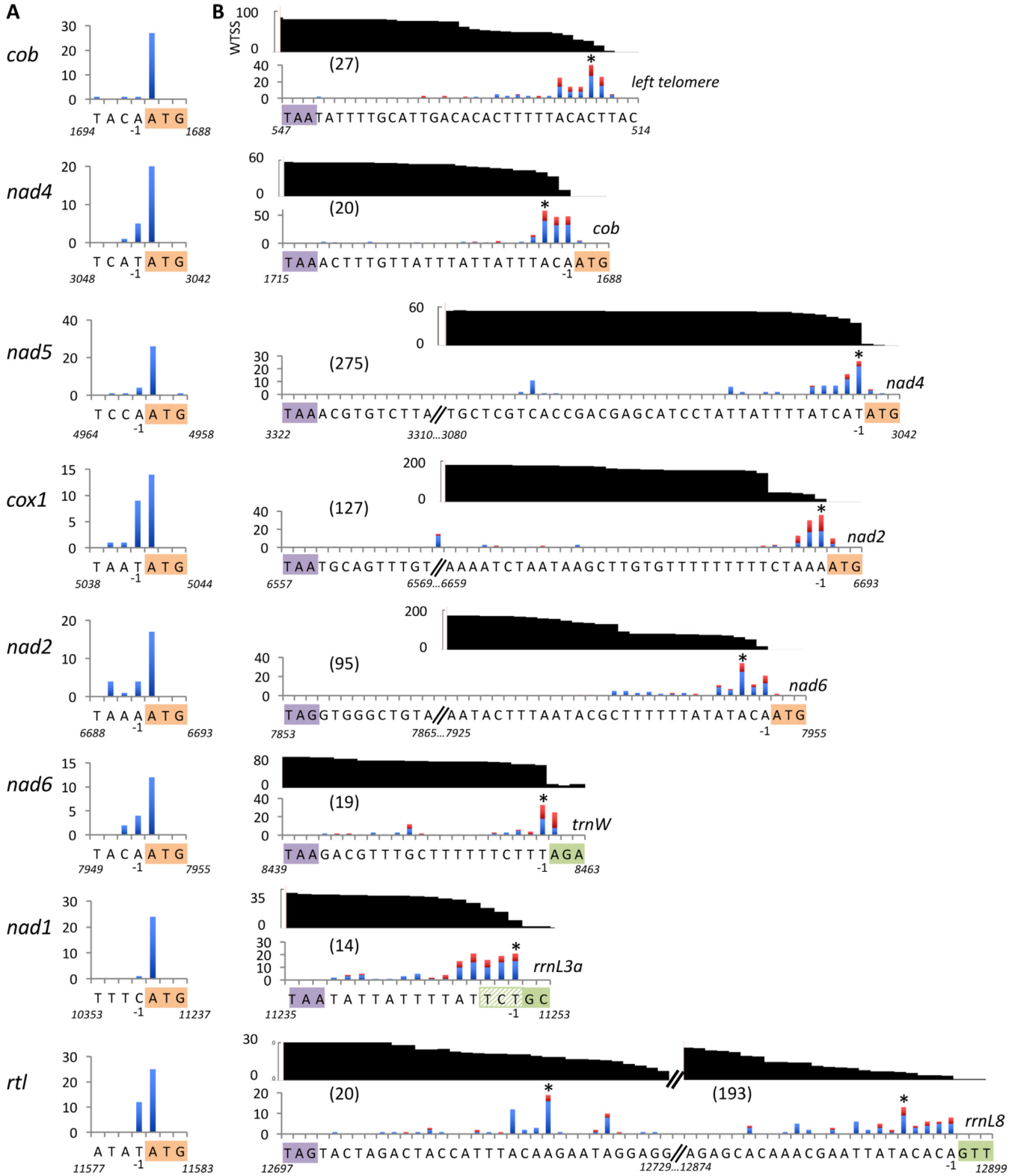

Figure 2. Mapping of the mRNAs extremities. (**A**) Number of 5′ RACE clones for each 5′ nucleotide position. The start codon is highlighted and its coordinate on the genome indicated. (**B**) Number of 3′ RACE and cRT-PCR clones (respectively blue and red bars) for each position of the 3′ nucleotide. The asterisk (*) indicates the main position(s) of the 3′-end, used to calculate the size of the 3′-UTR (numbers in brackets). The UTR sequence is shown under the *x*-axis, with the stop codon and the first nucleotides of the next gene highlighted in colors. For *rrnL3a,* two start positions (TC or GC) were observed in 5′ RACE and sRNA data (Supplementary Figure S2) respectively. The chart in black indicates the coverage, expressed as reads per million, of directional WTSS reads starting inside the gene of interest. For *cox1*, the abrupt drop at the end of the poly-U stretch is an artefact due to the presence in most reads of a larger number of U's, preventing mapping of the downstream nucleotides (see Supplementary Figure S3).

A

RNA	3′ RACE total	3′ RACE tail C-rich	3′ RACE tail A/U-rich	3′ RACE tail C:A/U-rich	3′ RACE <4	3′ RACE none	cRT-PCR total	cRT-PCR tail C-rich	cRT-PCR tail A/U-rich	cRT-PCR tail C:A/U-rich	cRT-PCR <4	cRT-PCR none
cob	110	32	9	1	26	42	66	4	4	1	18	39
nad4	146	48	14	1	40	43	61	18	4	0	16	23
nad5	108	25	11	0	21	51	24	3	4	0	3	14
cox1	79	37	11	2	11	18	63	14	15	1	8	25
nad2	95	7	10	1	29	48	32	4	2	1	15	10
nad6	94	23	10	2	18	42	66	6	14	0	18	28
nad1	90	17	7	2	23	41	36	8	5	0	11	12
rtl	155	6	26	5	36	82	38	2	8	0	5	23
Total	877	195	98	14	204	367	386	59	56	3	94	174

B

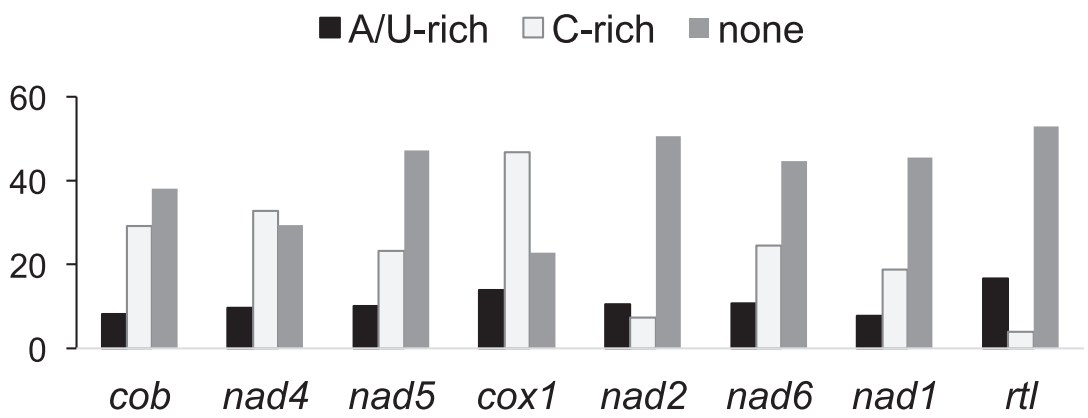

Figure 3. Summary of the 3′-UTR analysis of mRNAs. (**A**) For each method, the total number of sequences analyzed are indicated. Sequences were classified in three groups (see 'Materials and Methods' section): sequences without any added nt (none), sequences with 1–3 nt (<4) and sequences with a tail i.e. with more than 3 nt (tail). The tails were classified into three categories according to their nt content [(nt observed/nt tail length)*100]: tails with >50% of Cytidine (C-rich), tails with >50% of Adenosine/Uridine (A/U-rich) and tails with equal amounts of Cytidine and Adenosine/Uridine (C:A/U-rich). (**B**) Histogram summarizing the data obtained in (A) by the 3′ RACE approach and showing the percentage of transcripts with A/U-rich tail (black), C-rich tail (pale gray) and no tail (dark gray) for each mt gene.

C-rich tails	A/U-rich tails
UUUUAACCCCCCCCCCCCCC	UUUUAAAAAAAAAAAAAAAA
CCCCCCCCCCCCCCCCCCC	UUCUAAUAAACCC
UUAUUAAUUCCCCCCCCCC	AUAUUAAAAAA
CAUUUAAACCCCCCCCCA	UUAAAAACCCA
CCCCCCCUUUCCCCCCC	AAUAAAAAA
UUUCCCCCCCCCCCC	CCAUUAAAA
CCCCCCCCCCCC	AAACCCAAA
CACCCCCCCCCC	UAUACCC
CCCCCCCCCCA	AAAAAAA
CCCCCCCCU	AUUUUA
CCCCCCCCC	AUAAAA
CCCCCCCCC	AAAAA
CCCCCCCCU	UUAAA
AACCCCCAC	AUAAA
AACCCCCAC	CCAAA
AUCCCCCG	CCAAA
ACCCCCC	AAAAA
CCCCC	AGAC
CCCCC	AGAC
CUCC	AAUU

Figure 4. Randomly-selected C- (left) and A/U-rich (right) tails from 3′ RACE experiments.

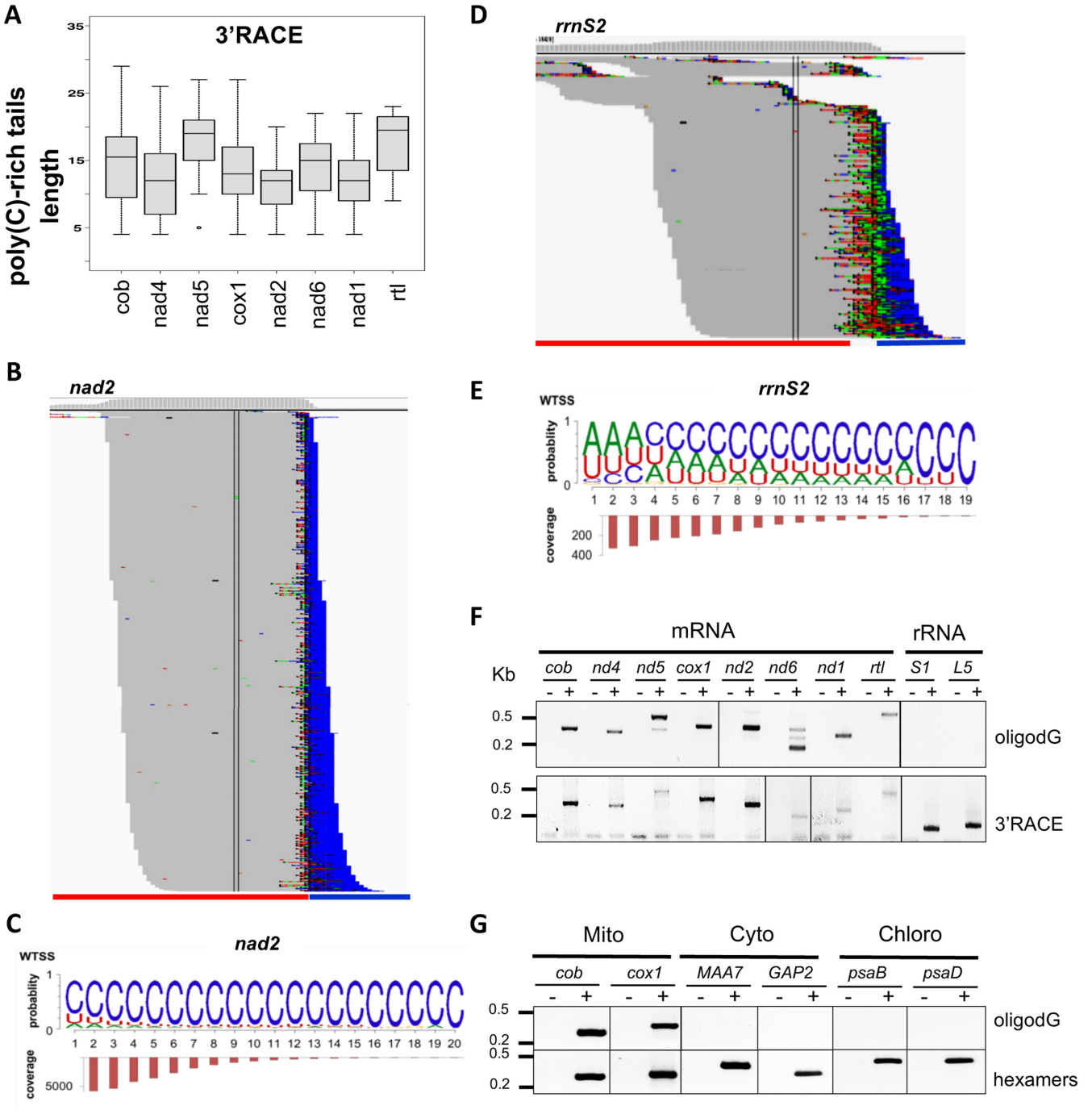

Figure 5. Analysis of C-rich tails. (**A**) Boxplot analysis of the length of C-rich tails observed by 3′ RACE. (**B**) and (**D**) 3′-tails at the end of *nad2* and *rrnS2*, respectively (red, with a blue line marking the next gene). Browser view of WTSS 3′-soft-clipped reads; coverage is plotted in log scale above the reads and only soft-clipped bases are colored (C: blue, U: red; A: green; G: brown). (**C**) and (**E**) the corresponding proportion-based sequence logos and length profiles for the 3′-tails using WTSS are shown. (**F**) RT-PCR to detect C-rich tails on mt transcripts by oligo(dG) priming. The 3′ RACE experiment serves as a control. (**G**) RT-PCR showing absence of C-rich tails on cytosolic and chloroplast transcripts. First strand cDNA synthesis was primed with oligo(dG) (top) or random hexamers (bottom).

the 3′-UTRs, respectively. They were short and showed no enrichment in C (Supplementary Figure S5B).

Comparison of mRNA and rRNA tails

In order to determine whether C-rich tails are specific of mRNAs or not, the 3′-ends of the 14 rRNA fragments were analyzed by 3′ RACE. Out of 304 sequences obtained, only 18 tails were found (6%), all short and A/U-rich, and in only 9 of the rRNA modules (Supplementary Figure S3). Directional WTSS confirmed the enrichment of rRNA tails in A/U compared to mRNAs, but some C nucleotides were also found, especially toward tail ends (Figure 5D, E and Supplementary Figure S4). These results were confirmed using a standard 20-cycle PCR protocol on oligo(dG)-primed cDNA: amplification was observed for all mRNAs, but not rRNAs (Figure 5F), in spite of the abundance of the templates. So far, polycytidylation has been reported neither for plastidial nor for cytosolic mRNAs. Using oligo(dG)-priming, the absence of C-rich tails in plastidial and cytosolic mRNAs of Chlamydomonas was confirmed for two transcripts of each compartment (Figure 5G). In conclusion, C-rich tails in Chlamydomonas are only present in mitochondria, where they are added preferentially to mRNAs.

Polycytidylation is a hallmark of the *Chlorophyceae*

In order to describe the phylogenetic realm of mt mRNA polycytidylation, 3′ RACE analyses were performed on total RNA from a number of species belonging to the three clades of Archaeplastida, i.e. eukaryotes with primary plastids. They include the Glaucocystophyte *C. paradoxa*, the red alga *C. crispus* and representatives from the green lineage including three Streptophytes (the moss *P. patens* and two flowering plants *A. thaliana* and *Solanum tuberosum*) and five Chlorophyte algae ranging from the basal *P. minor* to the non-photosynthetic *P. parva*, a close relative of *C. reinhardtii* (33). For each species, up to 3 mt transcripts were analyzed by 3′ RACE (Figure 6 and Supplementary Figure S6). Polycytidylation was found in two other Chlorophyceae, namely *P. parva* and *S. obliquus*. This trait was not found in the other species examined, in particular in the green algae *P. minor*, *O. tauri*, *Chlorella sorokiana* and *C. subellipsoidea*. It therefore appears that the molecular process leading to C-rich tails is specific to the Chlorophyceae class.

To further explore the Chlorophyceae, genome-wide WTSS datasets retrieved from the SRA or MMETSP databases were also analyzed (Supplementary Table S3). In the colonial alga *Volvox carteri*, complete intercistronic cleavage were found, mRNAs starting at AUG and C-rich tails at the end of the mRNAs, with rRNA tails enriched in As. The major difference with *C. reinhardtii* was that most 3′-UTR ended far upstream of the next gene. In *P. parva*, most mRNAs appeared to lack a 5′-UTR and showed C-rich tails, but the enrichment in A near the start was usually more pronounced than in *C. reinhardtii*, especially for mRNAs that were followed by a rRNA. In *C. leiostraca* and *C. moewusii*, two more distantly related *Chlamydomonas* species (34), polycytidylation of mRNAs was also found, even if the extent appeared lower than in *C. reinhardtii*.

DISCUSSION

The expression of the compact *C. reinhardtii* mt genome leads from two primary transcripts, to 3 tRNAs, 14 fragmented rRNA modules and 8 mRNAs containing no 5′-UTR and short 3′-UTR. If left and right telomeres are excluded, 905 nt (5.7%) are not part of a mature transcript. This organization is almost as compact as that of animal mt genomes, where a 'tRNA punctuation' model prevails for the generation of mature RNAs: primary transcripts are processed by endonucleolytic cleavage by RNase P and RNase Z respectively upstream and downstream of each tRNA to release mature rRNAs and mRNAs (35,36). In *C. reinhardtii* mitochondria as well, endonucleolytic cleavage is the major mode of RNA maturation. However, only one mRNA (*nad6*) has its 3′-end generated by PRORP, the enzyme that cleaves at the 5′ extremity of tRNAs in plants and algae (37) and no 5′-end coincides with the 3′-end of a tRNA, where RNase Z cleavage occurs. Thus, the identity of the endonuclease(s) cleaving at the AUG initiation codon of mRNAs and at the extremities of rRNAs in *Chlamydomonas* remains to be determined. In human mitochondria, FAST-Kinase Domain proteins have been shown to control, via their RAP domain, the maturation of transcripts not flanked by tRNAs (38). In Chlamydomonas, octotricopeptide repeat proteins with RAP domain such as those described in (39) may be involved, if targeted to the mitochondrion. A related question is how the cleavage sites are determined. Secondary structures within the 3′-UTR of mt mRNAs and palindromic repeats in the intergenic regions have been described (30,31) but they do not match the processing sites observed here. In addition, intervention of exonucleases is suggested not only by the 'spacer regions' not represented in mature RNAs (31) but also by the heterogeneity in length that we observe at the 3′-end of all mRNAs. The exonucleases identity is uncertain, because PNPase and the two RNase-II homologs in Chlamydomonas appear to function in the chloroplast (40,41).

The most intriguing feature of mRNA 3′-ends revealed by our study is the presence of non-templated post-transcriptionally added C-rich tails at the 3′-ends of mRNAs. In a study targeting all cell compartments, C was the only nucleotide not found to efficiently form tails (42). Thus, they potentially represent a novel way to regulate gene expression. Up to now, adenylation and uridylation were the two sole forms of nucleotide addition observed at the 3′-ends of mitochondrial RNAs. In animals, polyadenylation mostly stabilizes transcripts and completes the stop codon of several mRNAs, while in plants it serves as a transient signal for degradation (32,43). Accordingly, the 3 Streptophytes mt genes examined here by 3′-RACE showed only rare 1–2 nt additions, mostly As. In Myxomycetes, non-templated 3′ poly(U) tails were identified on edited mt mRNAs (44) while in *Trypanosoma brucei*, correctly edited mt transcripts are tailed with a poly(A/U) sequence of 200 to 300 nt which allows their recognition by the mt ribosomes and their translation activation (45).

C-rich tails were not observed in a previous study of Chlamydomonas mitochondria (46) which only retrieved A/U rich tails similar to those describe here, and long poly(A) tails (up to A_{32}) that were not observed in this

		mRNA	Sequences analyzed (3'RACE)	C-rich tails (%)
Glaucocystophyceae				
Cyanophoraceae	*Cyanophora paradoxa*	*nad4*	47	0
		cox1	61	0
Rhodophyceae				
Florideophyceae	*Chondrus crispus*	*cob*	51	0
Viridiplantae				
Streptophyta				
Bryophyta	*Physcomitrella patens*	*nad2*	69	0
Tracheophyta	*Arabidopsis thaliana*	*atp9*	68	0
	Solanum tuberosum	*atp9*	56	0
Chlorophyta				
Pedinophyceae	*Pedinomonas minor*	*cob*	50	0
		atp6	50	0
Prasinophyceae	*Ostreococcus tauri*	*nad4*	55	0
		cob	51	0
Trebouxiophyceae	*Coccomyxa subellipsoidea*	*nad4*	61	0
		cob	67	0
	Chlorella sorokiana	*nad4*	54	0
		atp9	53	0
		cox2	50	0
Chlorophyceae	***Sphaeropleales***			
	Scenedesmus obliquus	*cox1*	49	**53**
		atp9	66	**59**
	Chlamydomonales/Volvocales			
	Chlamydomonas reinhardtii	*All*	877	**22**
	Polytomella parva	*cob*	49	**49**
		cox1	52	**48**

Figure 6. Analyzed organisms on the Archaeplastida lineage show the presence of C-rich tails in *Chlorophyceae*.

study. This is probably due to the use of oligo(dT) priming, rather than to strain differences, as C-rich tails were found here in various *C. reinhardtii* strains as well as in other Chlorophycean algae belonging to diverse genera (*Chlamydomonas, Volvox, Polytomella*, *Scenedesmus*). Trebouxiophyceae and Prasinophyceae had only A-rich tails, while a red alga (*Chondrus*) showed no tail and the Glaucophyte *Cyanophora* showed an unexpected U-enrichment in the tails of *cox1*. The emergence of polycytidylation is therefore a late event in the algal lineage, coinciding approximately with rRNA fragmentation (47).

What are the mechanisms and consequences of polycytidylation in Chlamydomonas mitochondria? It is first notable that rRNAs showed short tails with very little poly(C) and almost never exactly at the mature 3′-end. This could be due to their rapid folding preventing access of the polycytidylation enzyme. mRNAs truncated within the CDS also showed rare and short tails, with no particular C-enrichment, which suggests that polycytidylation could be restricted to translatable mRNAs. It may be significant that *rtl*, the less conserved, less abundant and probably less translated mRNA, showed the weakest C-enrichment in its tails. In UTRs, C-rich tails are more frequent at or near the initial cleavage site, suggesting that polycytidylation is an early maturation event. UTRs that have undergone trimming can also be tailed, but apparently with much lower efficiency. If the putative 3′-5′ exonuclease(s) have a lower affinity for 3′-C residues, C-tails could constitute a protection, stabilizing the longer transcript, while decay would occur once the A/U rich part of the tail or the genome-encoded UTR is reached. It is also noteworthy that many tails start being A/U rich before C-additions prevail. This can be interpreted in the framework either of two different RNA-NTRs acting sequentially, or a single one changing specificity as polymerization progresses.

Since the lowly discriminating PNPase is believed to act mostly in the chloroplast (41), it is not a prime candidate for the generation of poly(C) tails. PAP1, a class-I NTR, is universally involved in cytosolic polyadenylation. The mt polycytidylation enzyme must therefore be sought among Class-II NTRs, even though the existence of still unknown classes of NTRs cannot be excluded. Among Chlamydomonas class-II NTRs (40), only PAP3 and PAP4 have been experimentally studied (46). As a first attempt to pinpoint the enzyme, we performed targeting predictions (Supplementary Table S4) and phylogenetic analysis (Figure 7) of Chlorophyta class-II NTRs, including a newly described isoform that we called PAP11. Among the 13 groups of NTRs that we identified, the majority was not predicted to reside in an organelle. This includes TRF4 and MUT68 which are known to act in the nucleolus or cytosol, respectively (48,49). Only PAP2, PAP3, PAP4, PAP4-L and PAP8 were consistently predicted to be organelle-targeted. Among them, PAP3 can be excluded as is has been shown to be a tRNA-CCA nucleotidyl transferase (46). PAP4-L, although related to Chlorophyceae-specific PAP4, is specific

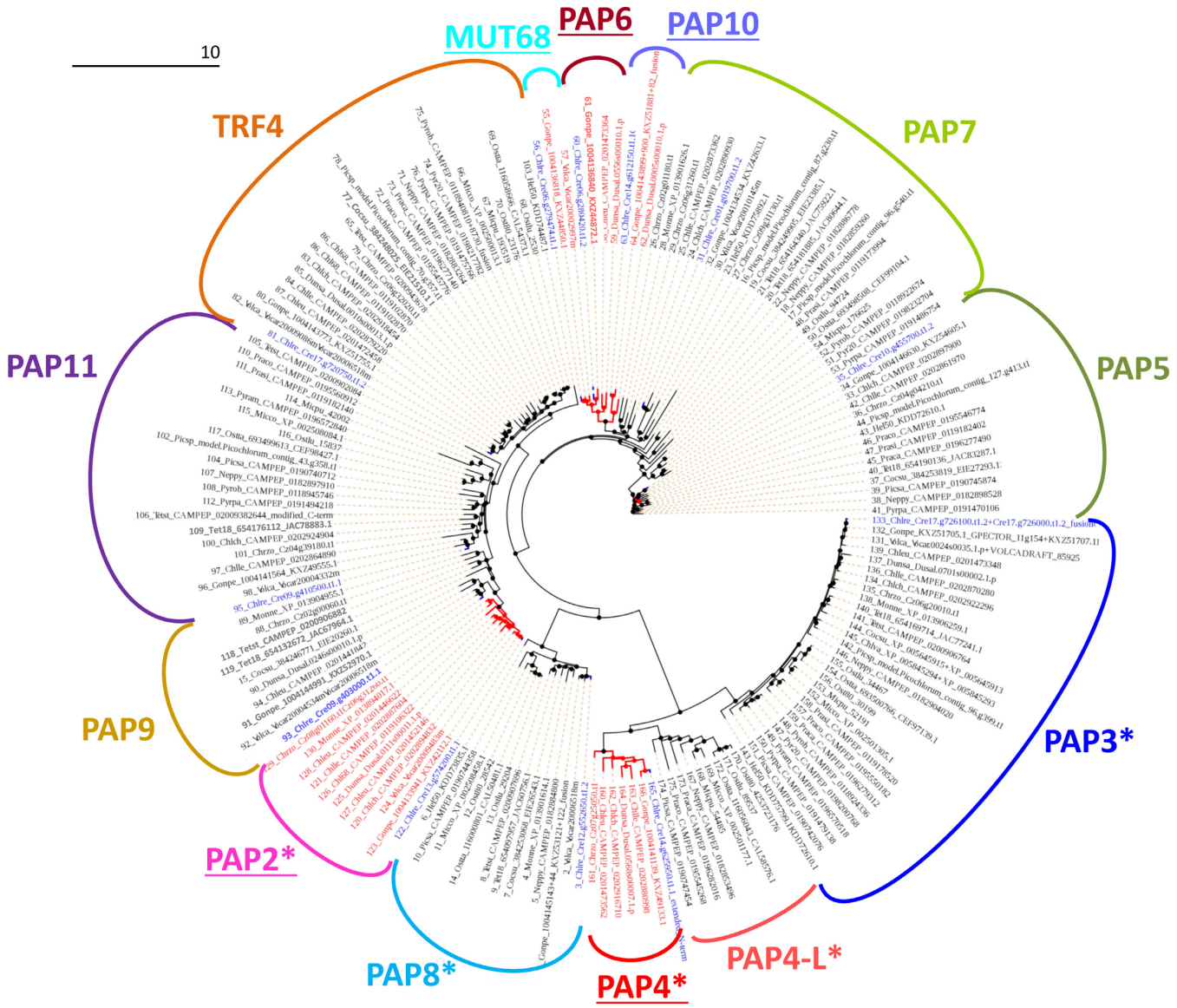

Figure 7. Phylogenetic tree of Chlorophyta class-II NTRs. Groups with Chlorophycean and non-Chlorphycean members are colored in red and black (branches and names), respectively. *Chlamydomonas reinhardtii* sequences are highlighted in blue. Groups are named based on the Chlamydomonas ortholog in Phytozome website (Chlamydomonas reinhardtii v5.5, http://phytozome.jgi.doe.gov/), those with only Chlorophycean members are underlined, those where most sequences are predicted organelle-targeted by TargetP or Predalgo are marked by an asterisk (Suplementary Table S4). Note that a PAP4 related family (PAP4-L) exists in *Prasinophyceae*. The abbreviations of the genus and species are listed in the Supplementary Table S4.

of Prasinophyceae which show A-rich tails. As to PAP8, it is present in all Chlorophytes and therefore not a prime candidate either. PAP4 and PAP2 thus appear as prime targets for future studies. The N-terminal sequence of CrPAP4 has been shown to address GFP to mitochondria in a heterologous system (46). This study also showed that CrPAP4 is able to elongate an A_{20} RNA substrate, but the nature of the added nucleotide(s) was not investigated. Among Chlamydomonas PAPs, CrPAP4 is the best hit of AGS1 which performs polyadenylation in Arabidopsis mitochondria (43). Note that the polymerization specificity of class-II NTRs can be low: *in vitro*, *Escherichia coli* poly(A) polymerase I can use all four triphosphate nucleotides with equal efficiency (50), which raises the question of whether the relative abundance of ATP and CTP plays a role in determining the composition of the tails. Further experiments will be required in order to identify the mitochondrial cytidyl transferase and shed light on the function of C-rich tails for Chlorophyceae mitochondrial gene expression processes.

DATA AVAILABILITY

The data as well as a novel annotation proposed here for the Chlamydomonas mt genome can be visualized on our genome browser (http://chlamy-organelles.ibpc.fr) and downloaded by right-clicking on the tracks.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank M. Gregorutti, M. Radoux and N. Cossemans for helpful technical assistance and J. Umen for sharing Volvox RNA-Seq data. A. Alioua is warmly acknowledged for sequencing work.

FUNDING

Centre National de la Recherche Scientifique (CNRS) in association with the Université de Strasbourg; Université Pierre et Marie Curie, Paris 06; Agence Nationale de la Recherche [ChloroRNP ANR-13-BSV7–0001-001]; 'Initiative d'Excellence' [DYNAMO ANR-11-LABX-0011–01, MITOCROSS ANR-11-LABX-0057]. The open access publication charge for this paper has been waived by Oxford University Press – *NAR* Editorial Board members are entitled to one free paper per year in recognition of their work on behalf of the journal.
Conflict of interest statement. None declared.

REFERENCES

1. Mai,N., Chrzanowska-Lightowlers,Z.M. and Lightowlers,R.N. (2017) The process of mammalian mitochondrial protein synthesis. *Cell Tissue Res.*, **367**, 5–20.
2. Ott,M., Amunts,A. and Brown,A. (2016) Organization and regulation of mitochondrial protein synthesis. *Annual Rev. Biochem.*, **85**, 77–101.
3. Salinas,T., Remacle,C. and Maréchal-Drouard,L. (2013) Mitochondrial translation in green algae and higher plants. In: Duchêne,AM (ed). *Translation in Mitochondria and Other Organelles*. Springer, Heidelberg; NY; Dordrecht; London, pp. 181–206.
4. Binder,S. and Brennicke,A. (2003) Gene expression in plant mitochondria: transcriptional and post-transcriptional control. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, **358**, 181–188.
5. Hammani,K. and Giegé,P. (2014) RNA metabolism in plant mitochondria. *Trends Plant Sci.*, **19**, 380–389.
6. Merchant,S.S., Prochnik,S.E., Vallon,O., Harris,E.H., Karpowicz,S.J., Witman,G.B., Terry,A., Salamov,A., Fritz-Laylin,L.K., Maréchal-Drouard,L. *et al.* (2007) The Chlamydomonas genome reveals the evolution of key animal and plant functions. *Science*, **318**, 245–250.
7. Remacle,C., Cardol,P., Coosemans,N., Gaisne,M. and Bonnefoy,N. (2006) High-efficiency biolistic transformation of Chlamydomonas mitochondria can be used to insert mutations in complex I genes. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 4771–4776.
8. Salinas,T., Larosa,V., Cardol,P., Maréchal-Drouard,L. and Remacle,C. (2013) Respiratory-deficient mutants of the unicellular green alga Chlamydomonas: A review. *Biochimie*, **100**, 207–218.
9. Grant,D. and Chiang,K.S. (1980) Physical mapping and characterization of Chlamydomonas mitochondrial DNA molecules: their unique ends, sequence homogeneity, and conservation. *Plasmid*, **4**, 82–96.
10. Boer,P.H. and Gray,M.W. (1988) Scrambled ribosomal RNA gene pieces in *Chlamydomonas reinhardtii* mitochondrial DNA. *Cell*, **55**, 399–411.
11. Duby,F., Cardol,P., Matagne,R.F. and Remacle,C. (2001) Structure of the telomeric ends of mt DNA, transcriptional analysis and complex I assembly in the dum24 mitochondrial mutant of *Chlamydomonas reinhardtii*. *Mol. Gen. Genet.*, **266**, 109–114.
12. Gray,M.W. and Boer,P.H. (1988) Organization and expression of algal (*Chlamydomonas reinhardtii*) mitochondrial DNA. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, **319**, 135–147.
13. Boer,P.H. and Gray,M.W. (1988) Genes encoding a subunit of respiratory NADH dehydrogenase (ND1) and a reverse transcriptase-like protein (RTL) are linked to ribosomal RNA gene pieces in *Chlamydomonas reinhardtii* mitochondrial DNA. *EMBO J.*, **7**, 3501–3508.
14. Ma,D.P., Yang,Y.W., King,T.Y. and Hasnain,S.E. (1990) The mitochondrial apocytochrome b gene from Chlamydomonas reinhardtii. *Plant Mol. Biol.*, **15**, 357–359.
15. Harris,E. (2001) Chlamydomonas as a model organism. *Annu. Rev. Plant Phys. Plant Mol. Biol.*, **52**, 363–406.
16. Stein,J. (1973) *Handbook of Phycological methods: Culture methods and growth measurements*. Cambridge University Press, Cambridge; London; NY; New Rochelle; Melbourne; Sydney.
17. Salinas,T., Duby,F., Larosa,V., Coosemans,N., Bonnefoy,N., Motte,P., Maréchal-Drouard,L. and Remacle,C. (2012) Co-evolution of mitochondrial tRNA import and codon usage determines translational efficiency in the green alga Chlamydomonas. *PLoS Genet.*, **8**, e1002946.
18. Loppes,R. and Radoux,M. (2001) Identification of short promoter regions involved in the transcriptional expression of the nitrate reductase gene in *Chlamydomonas reinhardtii*. *Plant Mol. Biol.*, **45**, 215–227.
19. Doublet,V., Ubrig,E., Alioua,A., Bouchon,D., Marcade,I. and Maréchal-Drouard,L. (2015) Large gene overlaps and tRNA processing in the compact mitochondrial genome of the crustacean *Armadillidium vulgare*. *RNA Biol.*, **12**, 1159–1168.
20. Cavaiuolo,M., Kuras,R., Wollman,FA., Choquet,Y. and Vallon,O. (2017) Small RNA profiling in Chlamydomonas: insights into chloroplast RNA metabolism. *Nucleic Acids Res.*, doi:10.1093/nar/gkx668.
21. Li,H. and Durbin,R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.
22. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
23. Barnett,D.W., Garrison,E.K., Quinlan,A.R., Stromberg,M.P. and Marth,G.T. (2011) BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics*, **27**, 1691–1692.
24. Makaroff,C.A., Apel,I.J. and Palmer,J.D. (1990) Characterization of radish mitochondrial atpA: influence of nuclear background on transcription of atpA-associated sequences and relationship with male-sterility. *Plant Mol. Biol.*, **15**, 735–746.
25. Quinlan,A.R. (2014) BEDTools: the Swiss-Army Tool for genome feature analysis. *Curr. Protoc. Bioinform.*, **47**, 11–34.
26. Criscuolo,A. (2011) morePhyML: improving the phylogenetic tree space exploration with PhyML 3. *Mol. Phylogenet. Evol.*, **61**, 944–948.
27. Emanuelsson,O., Brunak,S., von Heijne,G. and Nielsen,H. (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.*, **2**, 953–971.
28. Tardif,M., Atteia,A., Specht,M., Cogne,G., Rolland,N., Brugiere,S., Hippler,M., Ferro,M., Bruley,C., Peltier,G. *et al.* (2012) PredAlgo: a new subcellular localization prediction tool dedicated to green algae. *Mol. Biol. Evol.*, **29**, 3625–3639.
29. Ruwe,H., Wang,G., Gusewski,S. and Schmitz-Linneweber,C. (2016) Systematic analysis of plant mitochondrial and chloroplast small RNAs suggests organelle-specific mRNA stabilization mechanisms. *Nucleic Acids Res.*, **44**, 7406–7417.
30. Boer,P.H. and Gray,M.W. (1986) The URF 5 gene of *Chlamydomonas reinhardtii* mitochondria: DNA sequence and mode of transcription. *EMBO J.*, **5**, 21–28.
31. Boer,P.H. and Gray,M.W. (1991) Short dispersed repeats localized in spacer regions of *Chlamydomonas reinhardtii* mitochondrial DNA. *Curr. Genet.*, **19**, 309–312.
32. Lange,H., Sement,F.M., Canaday,J. and Gagliardi,D. (2009) Polyadenylation-assisted RNA degradation processes in plants. *Trends Plant Sci.*, **14**, 497–504.
33. Smith,D.R. and Lee,R.W. (2011) Nucleotide diversity of the colorless green alga *Polytomella parva* (Chlorophyceae, Chlorophyta): high for the mitochondrial telomeres, surprisingly low everywhere else. *J. Eukaryot. Microbiol.*, **58**, 471–473.
34. Pröschold,T., Harris,E.H. and Coleman,A.W. (2005) Portrait of a species: *Chlamydomonas reinhardtii*. *Genetics*, **170**, 1601–1610.
35. Ojala,D., Merkel,C., Gelfand,R. and Attardi,G. (1980) The tRNA genes punctuate the reading of genetic information in human mitochondrial DNA. *Cell*, **22**, 393–403.
36. Anderson,S., Bankier,A.T., Barrell,B.G., de Bruijn,M.H., Coulson,A.R., Drouin,J., Eperon,I.C., Nierlich,D.P., Roe,B.A.,

Sanger,F. *et al.* (1981) Sequence and organization of the human mitochondrial genome. *Nature*, **290**, 457–465.
37. Bonnard,G., Gobert,A., Arrive,M., Pinker,F., Salinas-Giegé,T. and Giegé,P. (2016) Transfer RNA maturation in Chlamydomonas mitochondria, chloroplast and the nucleus by a single RNase P protein. *Plant J.*, **87**, 270–280.
38. Boehm,E., Zaganelli,S., Maundrell,K., Jourdain,A.A., Thore,S. and Martinou,J.C. (2017) FASTKD1 and FASTKD4 have opposite effects on expression of specific mitochondrial RNAs, depending upon their endonuclease-like RAP domain. *Nucleic Acids Res.*, **45**, 6135–6146.
39. Boulouis,A., Drapier,D., Razafimanantsoa,H., Wostrikoff,K., Tourasse,N.J., Pascal,K., Girard-Bascou,J., Vallon,O., Wollman,F.A. and Choquet,Y. (2015) Spontaneous dominant mutations in chlamydomonas highlight ongoing evolution by gene diversification. *Plant Cell*, **27**, 984–1001.
40. Zimmer,S.L., Fei,Z. and Stern,D.B. (2008) Genome-based analysis of *Chlamydomonas reinhardtii* exoribonucleases and poly(A) polymerases predicts unexpected organellar and exosomal features. *Genetics*, **179**, 125–136.
41. Yehudai-Resheff,S., Zimmer,S.L., Komine,Y. and Stern,D.B. (2007) Integration of chloroplast nucleic acid metabolism into the phosphate deprivation response in Chlamydomonas reinhardtii. *Plant Cell*, **19**, 1023–1038.
42. Chang,H., Lim,J., Ha,M. and Kim,V.N. (2014) TAIL-seq: genome-wide determination of poly(A) tail length and 3′ end modifications. *Mol. Cell*, **53**, 1044–1052.
43. Hirayama,T., Matsuura,T., Ushiyama,S., Narusaka,M., Kurihara,Y., Yasuda,M., Ohtani,M., Seki,M., Demura,T., Nakashita,H. *et al.* (2013) A poly(A)-specific ribonuclease directly regulates the poly(A) status of mitochondrial mRNA in Arabidopsis. *Nat. Commun.*, **4**, 2247–2255.
44. Horton,T.L. and Landweber,L.F. (2000) Mitochondrial RNAs of myxomycetes terminate with non-encoded 3′ poly(U) tails. *Nucleic Acids Res.*, **28**, 4750–4754.
45. Aphasizhev,R., Suematsu,T., Zhang,L. and Aphasizheva,I. (2016) Constructive edge of uridylation-induced RNA degradation. *RNA Biol.*, **13**, 1078–1083.
46. Zimmer,S.L., Schein,A., Zipor,G., Stern,D.B. and Schuster,G. (2009) Polyadenylation in Arabidopsis and Chlamydomonas organelles: the input of nucleotidyltransferases, poly(A) polymerases and polynucleotide phosphorylase. *Plant J.*, **59**, 88–99.
47. Nedelcu,A.M., Lee,R.W., Lemieux,C., Gray,M.W. and Burger,G. (2000) The complete mitochondrial DNA sequence of *Scenedesmus obliquus* reflects an intermediate stage in the evolution of the green algal mitochondrial genome. *Genome Res.*, **10**, 819–831.
48. Sikorski,P.J., Zuber,H., Philippe,L., Sement,F.M., Canaday,J., Kufel,J., Gagliardi,D. and Lange,H. (2015) Distinct 18S rRNA precursors are targets of the exosome complex, the exoribonuclease RRP6L2 and the terminal nucleotidyltransferase TRL in *Arabidopsis thaliana*. *Plant J.*, **83**, 991–1004.
49. Ibrahim,F., Rymarquis,L.A., Kim,E.J., Becker,J., Balassa,E., Green,P.J. and Cerutti,H. (2010) Uridylation of mature miRNAs and siRNAs by the MUT68 nucleotidyltransferase promotes their degradation in Chlamydomonas. *Proc. Natl Acad. Sci. U.S.A.*, **107**, 3906–3911.
50. Yehudai-Resheff,S. and Schuster,G. (2000) Characterization of the *E.coli* poly(A) polymerase: nucleotide specificity, RNA-binding affinities and RNA structure dependence. *Nucleic Acids Res.*, **28**, 1139–1144.